UNITED STATES DISTRICT COURT

FOR THE DISTRICT OF MASSACHUSETTS

SINGULAR COMPUTING LLC, Plaintiff, v. GOOGLE LLC, Defendant.	Civil Action No. 1:19-cv-12551 FDS Hon. F. Dennis Saylor IV

DEFENDANT GOOGLE LLC'S MEMORANDUM OF LAW IN SUPPORT OF ITS MOTION FOR SUMMARY JUDGMENT OF NON-INFRINGEMENT

TABLE OF CONTENTS

TABLE OF AUTHORITIES

Page(s)

Federal Cases

Rules

I. INTRODUCTION[1]

Defendant Google LLC ("Google") respectfully seeks summary judgment of non-infringement, and submits that summary judgment is appropriate, because the accused Tensor Processing Unit ("TPU") boards do not and cannot satisfy a straightforward structural limitation of the Asserted Claims.[2]

Both Asserted Claims recite a structural limitation that requires a minimum number of "low precision high-dynamic range (LPHDR) execution units." In the context of the accused TPU boards, there is no dispute that this limitation requires at least 8,300 such units in order to establish infringement. Indeed, there is no dispute over *what* the physical components of the accused TPU boards are, how *many* there are, or how they function. But under the infringement theory of Plaintiff Singular Computing LLC's ("Singular") infringement expert, Dr. Sunil Khatri, there could only be, at most, 1,024 and 2,048 such LPHDR execution units in the accused version 2 and version 3 TPU boards, respectively. Summary judgment in Google's favor is therefore appropriate because of the simple mathematical fact that 1,024 and 2,048 are both less than 8,300. No amount of argument by Singular or opinions by Dr. Khatri can change that.

Because Singular does not (and cannot) contest the math, Dr. Khatri offers a workaround: according to him, the TPU boards contain hundreds of thousands more LPHDR execution units because they perform hundreds of thousands more of what he calls "LPHDR multiplication

[1] Throughout this brief, emphases have been added to quotations and internal quotation marks, alterations, footnotes, and citations omitted unless otherwise indicated, "SMF" refers to the accompanying Statement of Undisputed Material Facts in Support of Defendant Google LLC's Motion for Summary Judgment of Non-Infringement, and "Ex." citations refer to exhibits to the accompanying Declaration of Vishesh Narayen in Support of Defendant Google LLC's Motion for Summary Judgment of Non-Infringement ("Narayen Decl.").

[2] Singular asserts two claims (the "Asserted Claims"): dependent claim 53 of U.S. Patent No. 8,407,273 (the "'273 patent"), and dependent claim 7 of U.S. Patent No. 9,218,156 (the "'156 patent"). SMF ¶ 1.

operations" every clock cycle.[3] But the Asserted Claims are not method claims, and nowhere do they recite "clock cycles" or "LPHDR multiplication operations." Nor do the Asserted Claims define the required number of LPHDR execution units in terms of functionality or capability. Instead, the required number of LPHDR execution units is a ***structural*** limitation of the claims that, under Federal Circuit precedent, must turn on the ***structure*** of the TPU boards. Dr. Khatri's workaround is legally flawed because it turns on what a TPU allegedly *does* rather than what it *is*.

Because the undisputed facts establish that the accused TPU boards do not contain the required number of LPHDR execution units, the Court should grant summary judgment of non-infringement to Google.

In addition, in the event that the Court grants either (1) Google's pending motion to strike (Dkt. 408), or (2) the portion of the concurrently filed Motion to Exclude Certain Testimony of Sunil Khatri, Ph.D ("Khatri Daubert") seeking to exclude Dr. Khatri's improper interpretations of the Court's construction of the claimed "execution unit," which undergird his infringement analysis,[4] the Court can and should also grant summary judgment of non-infringement on the independent ground that Singular will be unable to carry its burden of proving infringement; Singular will be left without necessary expert opinion regarding alleged infringement. That absence of expert opinion would constitute a failure of proof that would also warrant summary judgment in Google's favor.

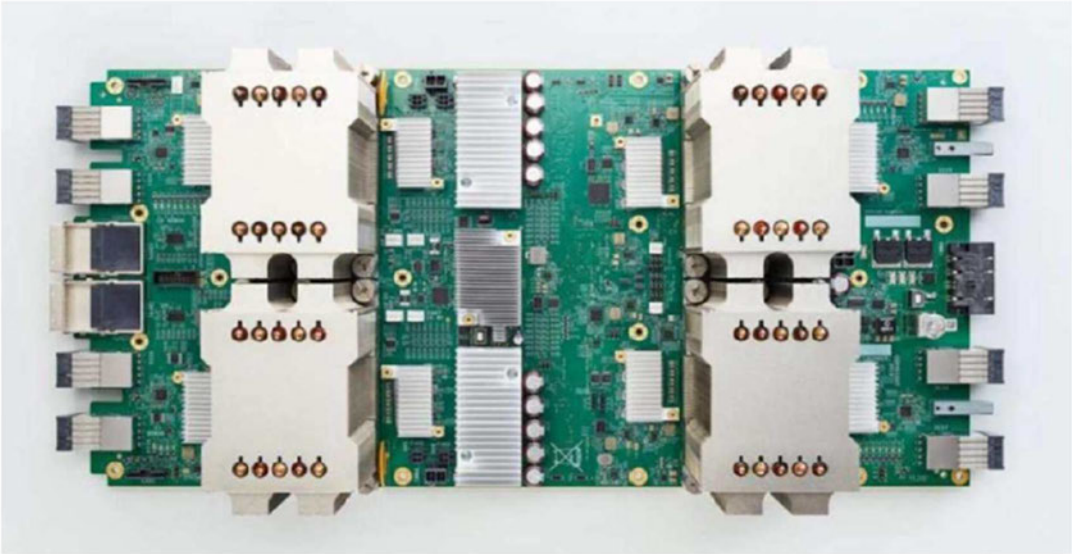
[3] A clock cycle is a measure of time that reflects the "pace" at which units within a computer operate. SMF ¶ 34.

[4] See Khatri Daubert § III.A (explaining why Dr. Khatri's infringement opinion impermissibly turns on his novel claim constructions). As other portions of the Khatri Daubert are not directly related to Singular's infringement theory, they are not relevant to the instant motion for summary judgment.

II. BACKGROUND

A. The Accused Tensor Processing Units

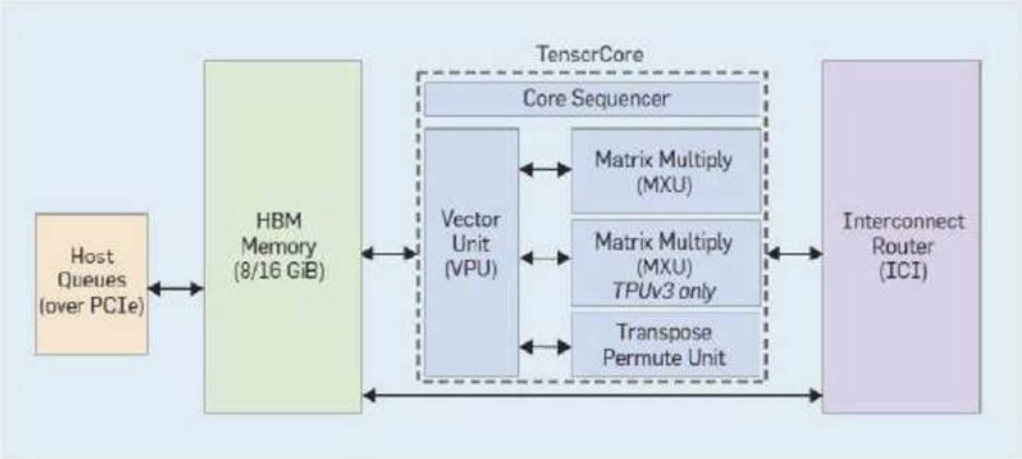
Google's TPU boards are application-specific integrated circuit boards designed to accelerate certain computing tasks. SMF ¶ 3. Singular contends that two versions of Google's TPU boards infringe: version 2 of the TPU board ("TPUv2") and version 3 of the TPU board ("TPUv3").[5] *Id.* ¶ 2. As an example, the image below shows a TPUv2 board. *Id.* ¶ 4.



As this motion centers on structural limitations of the Asserted Claims, it is important to count the TPU board's components, which speak to the TPU board's physical structure. Each TPUv2 board contains four integrated circuits (chips) known as "Jellyfish Chips" (JFCs) that are attached to the TPU board. *Id.* ¶ 5. Each JFC contains two "Tensor Cores," for a total of eight

[5] Notably, Singular's experts espouse inconsistent definitions of the accused product(s), as discussed in Google's concurrently filed Daubert motions as to Singular's experts Mr. Philip Green and Dr. Khatri. As explained in those motions, Google does not agree that the appropriate patent-practicing units (or the smallest salable patent-practicing units) are the TPUv2 and TPUv3 boards, even if Dr. Khatri's infringement theory were accepted. But for purposes of this motion—which does not depend on what the appropriate patent-practicing unit is—Google addresses this motion to the accused products as Singular and Dr. Khatri have identified them, i.e., the TPUv2 and TPUv3 boards.

Tensor Cores per TPUv2 board. *Id.* ¶ 6. The following block diagram illustrates one Tensor Core and some of its hardware components (in the dashed-line box). *Id.* ¶ 7.



As depicted above, each Tensor Core in the TPUv2 board contains several distinct components, including one "Matrix Multiply Unit" (MXU) capable of performing large matrix multiplication on 16-bit floating point values in the "bfloat16" (BF16) format. *Id.* ¶¶ 8–9. Thus, there is a total of eight MXUs on each TPUv2 board—1 MXU per Tensor Core × 2 Tensor Cores per JFC × 4 JFCs per TPUv2 board = 8 MXUs. *Id.* ¶ 8. As also depicted above, each of the eight Tensor Cores in the TPUv2 board also contains a "Vector Processing Unit" (VPU). *Id.* Among other things, the VPU contains 256 float conversion circuits ("rounding circuits"), each of which converts a 32-bit floating point value (FP32) to a 16-bit floating point value (BF16). *Id.* ¶ 10. Thus, there are 2,048 rounding circuits on each TPUv2 board—256 rounding circuits per VPU × 1 VPU per Tensor Core × 2 Tensor Cores per JFC × 4 JFCs per TPUv2 board = 2,048 rounding circuits. *Id.* Singular's expert, Dr. Khatri, refers to these rounding circuits as "precision-reducer circuits."[6] *Id.* ¶ 11.

[6] Although Dr. Khatri states that these rounding circuits ("precision-reducer circuits," as he refers to them) are in the MXU rather than the VPU, SMF ¶ 12, their specific location within the

The TPUv3 board is structurally similar to the TPUv2 board, but has twice as many MXUs and twice as many rounding circuits. *Id.* ¶¶ 13, 16–17. Specifically, each TPUv3 board comprises four integrated circuits known as "Dragonfish Chips" (DFCs). *Id.* ¶ 14. Like the JFCs in TPUv2, each DFC on the TPUv3 board contains two Tensor Cores, for a total of eight Tensor Cores per TPUv3 board. *Id.* ¶ 15. But each of the eight Tensor Cores in the TPUv3 board contains *two* MXUs (compared to the one MXU in each TPUv2 Tensor Core). *Id.* ¶ 16. There is therefore a total of sixteen MXUs on each TPUv3 board—2 MXUs per Tensor Core × 2 Tensor Cores per DFC × 4 DFCs per TPUv3 board = 16 MXUs. *Id.* The VPUs in each DFC each contain 512 rounding circuits—twice as many as in the TPUv2 VPUs. *Id.* ¶¶ 10, 17. Thus, there are 4,096 rounding circuits (Dr. Khatri's "precision-reducer circuits") on each TPUv3 board—512 rounding circuits per VPU × 1 VPU per Tensor Core × 2 Tensor Cores per DFC × 4 DFCs per TPUv3 board = 4,096 rounding circuits. *Id.* ¶ 17. The following chart summarizes the undisputed number of above-described components that are on each TPUv2 and TPUv3 board:

Component	**Total number on TPUv2 board**	**Total number on TPUv3 board**
Jellyfish Chips (TPUv2), or Dragonfish Chips (TPUv3)	4	4
Tensor Cores (two per JFC or DFC chip)	8	8
Matrix Multiply Units (MXUs) (one per Tensor Core in TPUv2, two per Tensor Core in TPUv3)	8	16

TPUs is not material to this Motion. In any event, there is no genuine dispute that the rounding circuits are located in the VPU. *Id.*

Vector Processing Units (VPUs) (two per JFC or DFC chip)	8	8
VPU rounding circuits (256 per VPU in TPUv2, 512 per VPU in TPUv3)	2,048	4,096

B. Singular's Infringement Theory

Singular accuses both the TPUv2 and TPUv3 boards of infringing the two Asserted Claims. SMF ¶ 2. Incorporating the limitations of the earlier claims from which they depend, the Asserted Claims read:

Claim 53 of the '273 patent:[7]

36. A device:

> comprising ***at least one first low precision high-dynamic range (LPHDR) execution unit*** adapted to execute a first operation on a first input signal representing a first numerical value to produce a first output signal representing a second numerical value,
>
> wherein the dynamic range of the possible valid inputs to the first operation is at least as wide as from 1/65,000 through 65,000 and for at least X=5% of the possible valid inputs to the first operation, the statistical mean, over repeated execution of the first operation on each specific input from the at least X % of the possible valid inputs to the first operation, of the numerical values represented by the first output signal of the LPHDR unit executing the first operation on that input differs by at least Y=0.05% from the result of an exact mathematical calculation of the first operation on the numerical values of that same input;
>
> wherein the number of LPHDR execution units in the device exceeds the non-negative integer number of execution units in the device adapted to execute at least the operation of multiplication on floating point numbers that are at least 32 bits wide.

43. The device of claim 36, ***wherein the number of LPHDR execution units in the device exceeds by at least one hundred the non-negative integer number of execution units in the device adapted to execute at least the operation of multiplication on floating point numbers that are at least 32 bits wide***.

[7] Claim 53 of the '273 patent depends from claim 43, which in turn depends from independent claim 36. SMF ¶ 19.

53. The device of claim 43, wherein the dynamic range of the possible valid inputs to the first operation is at least as wide as from 1/1,000,000 through 1,000,000.

SMF ¶ 19.

Claim 7 of the '156 patent:[8]

1. A device comprising:

> ***at least one first low precision high dynamic range (LPHDR) execution unit*** adapted to execute a first operation on a first input signal representing a first numerical value to produce a first output signal representing a second numerical value,
>
> > wherein the dynamic range of the possible valid inputs to the first operation is at least as wide as from 1/65,000 through 65,000 and for at least X=5% of the possible valid inputs to the first operation, the statistical mean, over repeated execution of the first operation on each specific input from the at least X% of the possible valid inputs to the first operation, of the numerical values represented by the first output signal of the LPHDR unit executing the first operation on that input differs by at least Y=0.05% from the result of an exact mathematical calculation of the first operation on the numerical values of that same input; and
>
> at least one first computing device adapted to control the operation of the at least one first LPHDR execution unit.

2. The device of claim 1, wherein the at least one first computing device comprises at least one of a central processing unit (CPU), a graphics processing unit (GPU), a field programmable gate array (FPGA), a microcode-based processor, a hardware sequencer, and a state machine.

3. The device of claim 2, ***wherein the number of LPHDR execution units in the device exceeds by at least one hundred the non-negative integer number of execution units in the device adapted to execute at least the operation of multiplication on floating point numbers that are at least 32 bits wide***.

7. The device of claim 3, wherein the dynamic range of the possible valid inputs to the first operation is at least as wide as from 1/1,000,000 through 1,000,000.

SMF ¶ 20.

[8] Claim 7 of the '156 patent depends from claim 3, which in turn depends from claim 2, which in turn depends from independent claim 1. SMF ¶ 20.

As shown above, the Asserted Claims are apparatus claims covering a "device" with particular structural attributes.[9] One key structural limitation in the Asserted Claims is the requirement of "at least one . . . low-precision high dynamic range (LPHDR) execution unit" (LPHDR EU). *Id.* ¶ 22. Another key structural limitation, which is recited identically in both Asserted Claims, requires that "the number of LPHDR execution units in the device *exceeds by at least one hundred* the non-negative integer number of execution units in the device adapted to execute at least the operation of multiplication on floating point numbers that are at least 32 bits wide." *Id.* ¶ 23. This "exceeds by at least one hundred" limitation thus imposes a numerosity requirement that entails comparing the number of [**A**] LPHDR EUs to the number of [**B**] execution units "adapted to execute at least the operation of multiplication on floating point numbers that are at least 32 bits wide."

To satisfy the numerosity requirement, the number of [**A**] (LPHDR EUs) in an accused device must "exceed[]" the number of [**B**] in the device "by at least one hundred."[10] If an accused device includes 500 of [**B**], for example, it must include at least 600 of [**A**] (LPHDR EUs) in order to satisfy this limitation, because 600 is the minimum number that "exceeds" 500 "by at least one hundred." If that accused device contains any fewer than 600 of [**A**], it does not meet this structural limitation and therefore cannot infringe. According to Singular, this numerosity requirement is important because "[s]uch heterogeneous architectures allow applications to make use of both traditional-precision and reduced-precision operations while retaining massive scale and parallelism in the form of a large number of LPHDR execution

[9] Dr. Khatri points to the TPUv2 and TPUv3 boards as the claimed "device." SMF ¶ 21.

[10] Singular's counting of the second type of execution units in the accused TPU boards is not disputed for purposes of this Motion. Therefore, in the interest of brevity, Google refers to this type of execution unit simply as [**B**].

units," which purportedly "enables new computer architectures that feature significant performance gains." Brief of Appellee, *Google LLC v. Singular Computing LLC*, No. 2022-1866, Dkt. No. 38, at 8 (Fed. Cir. Dec. 23, 2022).

Understanding Dr. Khatri's theory for what constitutes LPHDR EUs in the accused TPU boards, and how he counts them to satisfy the numerosity requirement, is critical to this motion. According to Dr. Khatri, each LPHDR EU in the accused TPU boards comprises: (i) two of the rounding circuits discussed above (Dr. Khatri's "precision-reducer circuits"); and (ii) a multiplication circuit in the MXU. SMF ¶ 24. For example, Dr. Khatri explicitly states that each LPHDR EU in the accused TPU boards "comprises the precision-reducer circuits that convert each of the FP32 input signals into low-precision BF16 signals . . . and the BF16 multiplication circuit that multiplies the BF16 signals and produces an output signal." Ex. 1 (Khatri Report) ¶ 140. Dr. Khatri created the following diagram in his report, purportedly based on deposition testimony and documents, to illustrate the components of what he posits are the LPHDR EUs in the accused TPU boards. SMF ¶ 24.

Dr. Khatri asserts that "[t]he above figure . . . shows the components of the LPHDR EU of the accused TPU devices. They include the [two] precision-reducer circuits (the boxes labeled

'R') . . . and the BF16 multiplication circuit (shown above as a circle marked with '✕')."[11] Ex. 1 ¶ 141. Dr. Khatri also opines that each LPHDR EU in the accused TPU boards is "pipelined" with "two pipeline stages." SMF ¶ 25. The first stage, according to Dr. Khatri, consists of the "two precision-reducer circuits" shown as blocks labeled '**R**' in his above diagram, which "convert[] two new FP32 input signals into BF16 signals." *Id.* ¶ 25 & n.22–23. The second pipeline stage, according to Dr. Khatri, comprises the BF16 multiplication circuit shown as the circle marked with "✕" in his diagram, which "multipl[ies] BF16 signals and produc[es] an output signal for the current LPHDR multiplication operation." *Id.* ¶ 25 & n.22–23. In short, according to Dr. Khatri's infringement theory, each LPHDR EU in Google's TPUs comprises two "precision-reducer circuits" and one "BF16 multiplier." *Id.* ¶¶ 24–25.

III. LEGAL STANDARD

Summary judgment is not just appropriate but mandatory if "the movant shows that there is no genuine dispute as to any material fact and the movant is entitled to judgment as a matter of law." Fed. R. Civ. P. 56(a). "Rule 56 [] mandates the entry of summary judgment 'against a party who fails to make a showing sufficient to establish the existence of an element essential to that party's case, and on which that party will bear the burden of proof at trial.'" *Coll v. PB Diagnostic Sys.*, 50 F.3d 1115, 1121 (1st Cir. 1995) (quoting *Celotex Corp. v. Catrett*, 477 U.S. 317, 322 (1986)).

"To establish literal infringement, every limitation set forth in a claim must be found in an accused product, exactly." *Duncan Parking Techs., Inc. v. IPS Grp., Inc.*, 914 F.3d 1347,

[11] To be clear, Google does not concede that Dr. Khatri's drawing fairly and accurately represents the structure of the TPU boards, the chips on the boards, or any part of the chips—particularly insofar as it incorrectly implies that the VPU's rounding circuits (Dr. Khatri's "precision-reducer circuits") are directly connected to MXU multiplier circuits.

1360 (Fed. Cir. 2019). "[A]ny deviation from the claim preclud[es] a finding of infringement." *Lantech, Inc. v. Keip Mach. Co.*, 32 F.3d 542, 547 (Fed. Cir. 1994).

Summary judgment of non-infringement is appropriate where "the patent owner's proof is deficient in meeting an essential part of the legal standard for infringement, because such failure will render all other facts immaterial." *TechSearch, L.L.C. v. Intel Corp.*, 286 F.3d 1360, 1369 (Fed. Cir. 2002). "Thus, an accused infringer 'is entitled to summary judgment, on the ground of non-infringement, by pointing out that the patentee failed to put forth evidence to support a finding that a limitation of the asserted claim was met by the structure in the accused devices.'" *Microsoft Corp. v. IQ Techs., Inc.*, 1 F.3d 1253 (Fed. Cir. 1993) (quoting *Johnston v. IVAC Corp.*, 885 F.2d 1574, 1578 (Fed. Cir. 1989)).

IV. ARGUMENT

A. The accused TPU boards do not meet the Asserted Claims' "exceeds by at least one hundred" limitations.

Because the Asserted Claims are apparatus claims, Singular must establish that the accused TPU boards "meet all of the structural limitations" in order to prove infringement. *Cross Med. Prods., Inc. v. Medtronic Sofamor Danket, Inc.*, 424 F.3d 1293, 1311–12 (Fed. Cir. 2005). The "exceeds by at least one hundred" limitation in both Asserted Claims imposes a numerosity requirement that is clearly structural: the number of [**A**] (LPHDR EUs) in the device must exceed the number of [**B**] in the device "by at least one hundred."[12]

[12] Structural limitations are the "physical structure and characteristics [of a device] specifically described in a claim." *Bicon, Inc. v. Straumann Co.*, 441 F.3d 945, 950 (Fed. Cir. 2006); *see also U.S. Filter Corp. v. Glegg Water Conditioning, Inc.*, 2005 WL 80947, at *1 (D. Mass. Jan. 13, 2005) ("[S]tructural limitations . . . define the physical characteristics of the device."). Limitations reciting a "plurality" of computing device components are structural limitations that require a minimum number of each component to be physically present. *See, e.g.*, *SIMO Holdings Inc. v. Hong Kong uCloudlink Network Tech. Ltd.*, 983 F.3d 1367, 1375–76 (Fed. Cir. 2021) ("Claim 8 is not a method claim but an apparatus claim. . . . that identifies **physical** components of the claimed **physical** device—beginning with 'a **plurality** of memory, processors, programs, communication circuitry, authentication data stored on a subscribed

1. The number of LPHDR EUs, if any, in the physical structure of the accused TPUs is less than the 8,300 needed to satisfy this limitation.

Singular cannot show that the numerosity requirement is satisfied because there are far fewer LPHDR EUs in the accused TPU boards than the Asserted Claims require, even under Dr. Khatri's theory for what constitutes a LPHDR EU.[13] Dr. Khatri opines—and Google does not dispute for purposes of this motion[14]—that the number of [**B**] on each accused TPU board (both TPUv2 and TPUv3) is 8,200. SMF ¶ 27. Specifically, Dr. Khatri concedes that each Tensor Core contains 1,025 of [**B**] because [redacted] *Id.* ¶ 28. Since each TPU board (both TPUv2 and TPUv3) contains eight Tensor Cores, Dr. Khatri concludes that the total number of [**B**] on each TPU board is 8,200 (1,025 of [**B**] in each Tensor Core × 8 Tensor Cores per TPU board = 8,200 of [**B**]). *Id.* ¶¶ 6, 15, 27–28. As Dr. Khatri states: "There are . . . 1,025 execution units in each Tensor Core that are adapted to perform 32-bit floating point multiplication. . . . This makes 8,200 execution units in each TPUv2 and TPUv3 board (1025 x 8) capable of 32-bit multiplication." Ex. 1 ¶¶ 234–235. Thus, to satisfy the "exceeds by at least one hundred"

identity module (SIM) card and/or in memory and non-local calls database.'" (emphasis in original)).

[13] To be clear, Google does not concede that the TPU boards or any component(s) within them comprise any LPHDR EUs or that the purported "operation" that Dr. Khatri identifies is a "first operation" within the meaning of the claims, but the Court need not reach those questions to resolve this Motion. This Motion is solely focused on the Asserted Claims' numerosity requirement, not Google's other non-infringement arguments pertaining to the claimed LPHDR EUs.

[14] As Google has previously informed the Court, in the pending appeals from the *inter partes* reviews of the '273 and '156 patents, Singular is advocating a claim construction before the Federal Circuit that, if adopted, would provide Google with further non-infringement defenses. *See, e.g.*, Dkt. 440. Specifically, as relevant here, the number of [**B**] in the TPUs would be dramatically higher under the position Singular is advocating to the Federal Circuit, providing an independent basis for non-infringement.

limitation under Singular's theory, each accused TPUv2 and TPUv3 board must contain at least **8,300** of [**A**]—that is, LPHDR EUs—because that is the minimum number that "exceeds [8,200] by at least one hundred."

But it is undisputed that there are only 2,048 rounding circuits on each TPUv2 board and only 4,096 rounding circuits on each TPUv3 board. SMF ¶¶ 10, 17, 30–31. And under Dr. Khatri's infringement theory, each LPHDR EU in the TPUs comprises two rounding circuits. *Id.* ¶ 24. Thus, there could only be, at most, 1,024 of [**A**] on each TPUv2 board (2,048 ÷ 2) and 2,048 on each TPUv3 board (4,096 ÷ 2). Both 1,024 and 2,048 are far less than the 8,300 needed to satisfy the "exceeds by at least one hundred" limitation, which warrants summary judgment of no infringement as a matter of simple math.

2. Dr. Khatri's workaround for the "exceeds by at least one hundred" limitation is legally flawed because it impermissibly bases purported infringement on what the TPU allegedly does rather than what it is.

Because the number of rounding circuits is plainly insufficient to satisfy the numerosity requirement for LPHDR EUs, Dr. Khatri tenders a workaround that misrepresents, as a matter of law, the number of LPHDR EUs in the TPU boards based on how many purported *operations* the TPU boards can perform each clock cycle rather than what actually exists in their physical structure. This theory cannot create any dispute of material fact concerning infringement because it is legally flawed, basing infringement of a limitation that recites a certain *number* of a physical structure on what the structure ***does***, rather than on the number of instances of that physical structure that exist on the device.

As explained above, Dr. Khatri opines that "the ***components*** of the [purported] LPHDR EU of the accused TPU devices . . . include the [two] precision-reducer circuits . . . and the BF16 multiplication circuit." Ex. 1 ¶ 141. Despite the undisputed fact that there are only 2,048 rounding circuits on each TPUv2 board and only 4,096 rounding circuits on each TPUv3 board

(SMF ¶¶ 10, 17, 30–31), Dr. Khatri opines that there are ***131,072*** LPHDR EUs in the TPUv2 board and ***262,144*** LPHDR EUs in the TPUv3 board because the TPU boards perform that many of what he calls "LPHDR multiplication operations" each clock cycle. *Id.* ¶¶ 32–33. According to Dr. Khatri: "Every clock cycle, MXUs of the accused TPUv2/3 devices complete 16,384 . . . LPHDR multiplication operations . . . . Thus, there [are] 16,384 LPHDR execution units per MXU." Ex. 1 ¶¶ 224–225. Likewise, Dr. Khatri opines: "[E]ach MXU . . . completes 16,384 floating-point operations—*i.e.*, LPHDR multiplications—per cycle, and thus contains 16,384 independent LPHDR execution units[.]" *Id.* ¶ 228. Since each TPUv2 board contains 8 MXUs, and each TPUv3 board contains 16 MXUs, Dr. Khatri extrapolates that there must be 131,072 LPHDR EUs in the TPUv2 board (16,384 × 8) and 262,144 in the TPUv3 board (16,384 × 16). SMF ¶¶ 8, 16, 32–33. Dr. Khatri purports to justify counting operations in lieu of physical structures because "computer engineers quantify computational efficiency by counting the number of completed executions per unit time." *Id.* ¶ 35; Ex. 1 ¶ 233.

This is the crux of Dr. Khatri's workaround: counting the operations purportedly performed by the TPU boards in lieu of the number of distinct (purported) LPHDR EUs that are physically present in the *structure* of the TPU boards. But the Asserted Claims are apparatus claims, not method claims, and the Federal Circuit has repeatedly made clear that "apparatus claims cover what a device *is*, not what a device *does*." *Hewlett-Packard Co. v. Bausch & Lomb, Inc.*, 909 F.2d 1464, 1468 (Fed. Cir. 1990) (emphasis in original). As the Federal Circuit has also explained, it is therefore "improper to construe non-functional claim terms in apparatus claims in a way that makes infringement or validity turn on the way an apparatus is later put to

use,"[15] *Edgewell Pers. Care Brands, LLC v. Munchkin, Inc.*, 998 F.3d 917, 921 (Fed. Cir. 2021)—which is exactly what Dr. Khatri does by counting the number of operations purportedly performed per clock cycle by the TPU boards when put to use, in lieu of the number of (purported) LPHDR EUs that are physically present in the structure of the TPUs.

Because infringement of apparatus claims properly turns on what the accused product *is* rather than what it *does*, district courts have granted, and the Federal Circuit has affirmed, summary judgment of non-infringement when plaintiffs improperly use an accused product's operation to try to prove infringement of structural limitations. For example, in *Reckitt Benckiser LLC. v. Aurobindo Pharma Ltd.*, the limitations at issue required that the claimed drug compound "comprise two distinct formulations" of an active ingredient: an immediate-release and a sustained-release formulation. 239 F. Supp. 3d 822, 828 (D. Del. 2017), *aff'd*, 737 Fed. App'x 537 (Fed. Cir. 2018). As the court explained, these were "structural limitations about two distinct formulations." *Id.* at 833. But rather than "provide evidence regarding the physical structure of Aurobindo's product to demonstrate that it includes more than a single formulation," Reckitt's infringement theory was "based entirely on how Aurobindo's [drug] product performs"—namely that it "behaves as if it contains two distinct formulations, and so [it] must have two formulations." *Id.* at 828-30. In granting summary judgment of non-infringement,

[15] The rationale for this is simple: "Construing a non-functional term in an apparatus claim in a way that makes direct infringement turn on the use to which an accused apparatus is later put confuses rather than clarifies, frustrates the ability of both the patentee and potential infringers to ascertain the propriety of particular activities, and is inconsistent with the notice function central to the patent system." *Paragon Sols., LLC v. Timex Corp.*, 566 F.3d 1075, 1091 (Fed. Cir. 2009); *see also In re Swinehart*, 439 F.2d 210, 213 (C.C.P.A. 1971) ("Functional terminology may render a claim quite broad. By its own literal terms a[n apparatus] claim employing such language covers any and all embodiments which perform the recited function. Legitimate concern often properly exists, therefore, as to whether the scope of protection defined thereby is warranted . . . .").

which was affirmed by the Federal Circuit, the court explained that Reckitt's evidence of how Aurobindo's accused drug performs "do[es] not speak to the formulation of the product" itself. *Id.* at 833. Likewise, here, Singular cannot establish infringement by counting the number of operations (purportedly) performed by the TPU boards—*i.e.*, Dr. Khatri's so-called "LPHDR multiplication operations"—in lieu of the number of (purported) LPHDR EUs physically present in the structure of the TPU boards.

Apparatus claims sometimes recite structural limitations in terms of the apparatus's capabilities when put into operation, but the numerosity requirement in the Asserted Claims does not do so. The numerosity requirement is not, for example, claimed as "a plurality of execution units capable of performing 8,300 LPHDR multiplication operations per clock cycle." Indeed, the Asserted Claims nowhere recite "LPHDR multiplication operations" or "clock cycles" (or any other unit of time). "Inventors are masters of their claims, and the words they use to describe and claim their invention are decisive and binding." *Bio-Rad Lab'ys, Inc. v. Int'l Trade Comm'n*, 998 F.3d 1320, 1331 (Fed. Cir. 2021). As the master of its claims, Singular chose to incorporate a structural limitation into the Asserted Claims that requires a minimum number of LPHDR EUs to be physically present; the question of infringement must therefore turn on the *structure* of the TPU boards, not how many (or few) operations they can perform when put to use. *See Agrizap, Inc. v. Woodstream Corp.*, 514 F. Supp. 2d 714, 719 (E.D. Pa. 2007) ("Claim 16 is solely an apparatus claim, so it has nothing to do with functionality. . . . Therefore, any arguments about the code and the microprocessor carrying out the same function as Claim 16 are irrelevant."), *rev'd on other grounds*, 520 F.3d 1337 (Fed. Cir. 2008).

Lutron Elecs. Co. v. Crestron Elecs., Inc., 970 F. Supp. 2d 1229 (D. Utah 2013), illustrates, under analogous facts, the difference between claim limitations defined by structure

and those defined by capability. In *Lutron*, the claims required "at least one control device coupled to the electronic device by a wire connection for providing power to the electrical device," which Crestron argued was not met because it was undisputed that the control device in its accused products was not "coupled . . . by a wire" to an electrical device. *Id.* at 1233. Lutron argued that the "coupled . . . by a wire" limitation "merely informs how to carry out the purpose of providing power, rather than requiring actual coupling before direct infringement may occur." *Id.* at 1235. The court disagreed, holding that the coupled-by-a-wire requirement was "a structural limitation" that could not be satisfied merely by pointing to capability, and, therefore, "direct infringement . . . cannot occur until the control device actually is coupled, by wire, to the electrical device." *Id.* at 1235-36. The court observed that "if the relevant claim limitation [] were 'at least one control device . . . for providing power to the electrical device,' the court would agree that [] that phrase recites capability rather than a structural limitation." *Id.* at 1235.

Dr. Khatri's workaround theory analogously hinges on a functional requirement not actually recited in the Asserted Claims—*i.e.*, the number of "LPHDR multiplication operations" purportedly performed every clock cycle—to satisfy the requirement that the TPU boards contain the claimed *number* of LPHDR execution units, which is a structural limitation. But this theory is flawed as a matter of law, and Singular cannot "modify the scope of its claim" by substituting a functional requirement for a structural one. *Id.* at 1236.

To more plainly illustrate why Dr. Khatri's workaround is impermissible, consider a hypothetical accused device that undisputedly contained 8,200 of [**B**] and 8,300 distinct, physical LPHDR EUs (i.e., [**A**]), but was designed in such a way that it performed only 4,150 of Dr. Khatri's "LPHDR multiplication operations" every clock cycle. Such a device would satisfy the Asserted Claims' numerosity requirement despite the fact that the number of "LPHDR

multiplication operations" it performs every clock cycle (4,150) does not "exceed[] [8,200] by at least one hundred," precisely because the numerosity requirement is a structural limitation that does not "turn on the way [the device] is later put to use." *Edgewell*, 998 F.3d at 921.

B. Summary judgment is warranted if the Court grants Google's pending motion to strike or the relevant portion of the Khatri Daubert because Singular cannot show infringement.

If the Court grants either Google's pending motion to strike (Dkt. 408), or the portion of the Khatri Daubert seeking to exclude the bases for Dr. Khatri's infringement opinion, Singular cannot meet its burden of demonstrating infringement of the Asserted Claims because it will be left without necessary expert opinion on how Google's TPUs purportedly comprise any LPHDR EUs. Singular's inability to provide sufficient evidence of infringement in that scenario would be an independent basis for summary judgment.

As discussed in the memorandum accompanying Google's motion to strike, the Court should strike the paragraphs and exhibits of Dr. Khatri's infringement report that, for the first time, disclosed a theory of infringement based on a purported two-stage "LPHDR multiplication operation," because this theory was never previously disclosed in Singular's preliminary or supplemental invalidity contentions and constituted a unjustifiable, tactical shift in Singular's infringement theory.[16] *See generally* Dkt. 409.

Furthermore, the Khatri Daubert explains how Dr. Khatri's mapping of the claim elements to the accused TPU boards hinges on Dr. Khatri's novel (and incorrect) interpretations of the Court's construction of the claimed "execution unit." *See, e.g.*, Khatri Daubert § III.A. In other words, Dr. Khatri impermissibly performed claim construction on the Court's claim

[16] Google specifically requested that the Court strike at least paragraphs 128–133, 140–142, 144–146, 148, 158–174, 178–210, 212, 222–223, 228–233, 235, 237–238, and Exhibits D–G of Dr. Khatri's infringement report. Dkt. 409 at 11.

construction. *Id.* If the Court excludes these opinions, then Singular would be bereft of expert opinion on how and why the accused TPU boards purportedly contain LPHDR EUs.

Accordingly, if the Court either (1) strikes the paragraphs of Dr. Khatri's report (or corresponding exhibits) espousing a previously undisclosed infringement theory or (2) excludes his opinions that construe the Court's construction of "execution unit," Singular will have no expert opinion on how the TPU boards purportedly satisfy key claim limitations relating to LPHDR EUs, which would mandate summary judgment.[17] *See, e.g.*, *Centricut, LLC v. Esab Group, Inc.*, 390 F.3d 1361, 1370 (Fed. Cir. 2004) ("[I]n a case involving complex technology, where the accused infringer offers expert testimony negating infringement, the patentee cannot satisfy its burden of proof by relying only on testimony from those who are admittedly not expert in the field. That is what happened here, and the patentee thus failed to satisfy its burden of proof."); *Midwest Athletics & Sports All. LLC v. Xerox Corp.*, 2022 WL 4493007, at *36 (W.D.N.Y. Sept. 28, 2022) ("Here, without the stricken portions of the Mitzenmacher Report, Plaintiff has identified no evidence from which a jury could find that the accused products satisfy the 'input device' limitations of claim 1 of the '314 Patent and claim 2 of the '974 Patent or the 'selectively associating' limitation of claim 51 of the '314 Patent. Accordingly, Defendant has demonstrated that it is entitled to summary judgment of non-infringement on these claims."); *Anticancer, Inc. v. Cambridge Rsch. & Instrumentation, Inc.*, 2009 WL 9115821, at *4 (S.D. Cal. Feb. 13, 2009) ("Summary judgment is appropriate against a party who bases its opposition on theories omitted from its infringement contentions." (collecting cases)).

[17] Singular could not fill the gap left by the to-be-stricken-or-excluded portions of Dr. Khatri's report with fact or lay opinion testimony because the accused technology in this case is undisputedly complex. *See, e.g.*, *AquaTex Indus., Inc. v. Techniche Sols.*, 479 F.3d 1320, 1329 n.7 (Fed. Cir. 2007) ("[E]xpert infringement testimony is generally required in cases involving complex technology.").

Accordingly, if the Court grants Google's motion to strike or excludes Dr. Khatri's opinions founded on an incorrect reading of the Court's claim construction, the Court can and should grant summary judgment of non-infringement on this additional, independent basis.

V. CONCLUSION

Summary judgment of non-infringement is appropriate because the undisputed facts establish that the accused TPU boards do not and cannot satisfy the Asserted Claims' "exceeds by at least one hundred" limitation. Summary judgment is independently warranted if the Court grants Google's pending motion to strike (Dkt. 408) or excludes Dr. Khatri's incorrect interpretation and application of the Court's claim construction to the accused TPU boards, either of which would leave Singular without necessary expert opinion that the TPU boards comprise any LPHDR EUs. Google respectfully requests that the Court grant summary judgment of non-infringement on both grounds.

Respectfully submitted,

Dated: April 28, 2023

By: */s/ Nathan R. Speed*

GREGORY F. CORBETT (BBO #646394)
gcorbett@wolfgreenfield.com
NATHAN R. SPEED (BBO #670249)
nspeed@wolfgreenfield.com
ELIZABETH A. DiMARCO (BBO #681921)
edimarco@wolfgreenfield.com
ANANT K. SARASWAT (BBO #676048)
asaraswat@wolfgreenfield.com
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, MA 02210
Telephone: (617) 646-8000
Fax: (617) 646-8646

ROBERT VAN NEST (pro hac vice)
rvannest@keker.com
MICHELLE YBARRA (pro hac vice)
mybarra@keker.com
ANDREW BRUNS (pro hac vice)
abruns@keker.com
VISHESH NARAYEN (pro hac vice)
vnarayen@keker.com
CHRISTOPHER S. SUN (pro hac vice)
csun@keker.com
ANNA PORTO (pro hac vice)
aporto@keker.com
DEEVA SHAH (pro hac vice)
dshah@keker.com
STEPHANIE J. GOLDBERG (pro hac vice)
sgoldberg@keker.com
KEKER, VAN NEST & PETERS LLP
633 Battery Street
San Francisco, CA 94111-1809
(415) 391-5400

MICHAEL S. KWUN (*pro hac vice*)
mkwun@kblfirm.com
ASIM BHANSALI (*pro hac vice*)
abhansali@kblfirm.com
KWUN BHANSALI LAZARUS LLP
555 Montgomery Street, Suite 750
San Francisco, CA 94111
(415) 630-2350

Matthias A. Kamber (*pro hac vice*)
matthiaskamber@paulhastings.com
PAUL HASTINGS, LLP
101 California Street
Forty-Eighth Floor
San Francisco, CA 94111
(415) 856-7000

Counsel for Defendant Google LLC

CERTIFICATE OF SERVICE

I certify that this document is being filed through the Court's electronic filing system, which serves counsel for other parties who are registered participants as identified on the Notice of Electronic Filing (NEF). Any counsel for other parties who are not registered participants are being served by first class mail on the date of electronic filing.

/s/ *Nathan R. Speed*
Nathan R. Speed